

TÉCNICAS ESTADÍSTICAS APLICADAS AL ANÁLISIS DE RENDIMIENTO EN EL DEPORTE

Por

Sergio Herrera Burró

TRABAJO FIN DE GRADO
PRESENTADO PARA OPTAR
AL GRADO EN ESTADÍSTICA
POR LA UNIVERSIDAD DE SEVILLA

Sergio Herrera Burró

UNIVERSIDAD DE SEVILLA

Fecha: **Junio 2018**

Autor: **Sergio Herrera Burró**

Título: **Técnicas Estadísticas Aplicadas al Análisis de
Rendimiento en el Deporte**

Dpto: **Estadística e Investigación Operativa**

Firma del autor

SE PROHIBE LA REPRODUCCIÓN TOTAL O PARCIAL DE ESTE
TRABAJO SIN EL PERMISO EXPRESO DEL AUTOR O DE LA UNIVERSIDAD
DE SEVILLA.

Resumen	I
Abstract	II
1. INTRODUCCION	0
1.1. Introducción	0
2. ANALISIS DE REDES	2
2.1. Teoría del análisis de redes	2
2.1.1. Centralidad del grado	5
2.1.2. Intermediación	7
2.1.3. Cercanía	9
2.1.4. Pagerank	11
2.1.5. Clustering y comunalidades	12
3. APLICACION DE REDES EN FUTBOL	14
3.1. Aplicaciones del análisis de redes en el fútbol	14
3.1.1. Caso I: Previa a la final de Sudáfrica. España vs Holanda	15
3.1.2. Caso II: Previa al 3º puesto de Sudáfrica. Alemania vs Uruguay .	18
4. APLICACION A LA FINAL FIFA	20
4.1. Análisis de redes en RStudio	20
4.1.1. Paquete iGraph	20
4.1.2. Resultados	21
Anexos	26
A. CODIGOS DE R	26

En este trabajo presentaremos algunas de las diferentes herramientas del análisis de redes para describir la estrategia de los equipos de fútbol. Utilizando los datos de pases de los equipos puestos a disposición por la FIFA, durante la Copa Mundial de los años 2010 y 2014, construyendo para cada equipo una red ponderada de pase donde los nodos son los jugadores y los arcos son los pases.

La red de pase ofrece una inspección visual directa de la estrategia del equipo, en la cual podemos identificar el patrón del juego o encontrar las debilidades potenciales. Usando diferentes medidas de centralidad el análisis puede ser más profundo y nos permitirá obtener valores que determinan la importancia de cada jugador en el juego, la popularidad de un jugador y el efecto de alejar a los jugadores del juego.

In this final degree project are presented some of the different tools of network theory to describe the strategy of football teams. Using the pass data of the teams made available by FIFA, during the Word Cup of the years 2010 and 2014. Weighted pass network for each team is propose. Furthermore, in which the nodes are the players and the ties are the passes.

The pass network offers a direct visual inspection of the team's strategy, in which we can identify the pattern of the game or find potential weaknesses. But using different measures of centrality analysis can be deeper and allow us to obtain values that determine the imprtance of each player in the game, the popularity of a player and the effect of keeping players away from the game.

1 INTRODUCCION

“Si tú tienes el balón, el rival no lo tiene”

Johan Cruyff (1947-2016)

1.1. Introducción

En el análisis de los encuentros de fútbol es habitual encontrar referencias a recuperaciones, pérdidas de balón, centros al área, tiros a puerta y otras acciones individuales desarrolladas por los jugadores durante el partido, ver Castellano y Blanco (2002).

Este uso tradicional de las estadísticas individuales, centradas en aspectos técnicos y finales, presenta limitaciones cuando pretenden evaluar el componente táctico del deporte, ver Chicote.J, Morante.J.C, y Vaquera.A (2009), ya que el comportamiento de los jugadores en el campo depende de relaciones de cooperación-oposición, de forma en la que la acción individual viene en parte determinada por las sucesivas configuraciones tácticas adoptadas por el conjunto de jugadores en interacción, ver Lago.C y Anguera.M.T (2003).

Una propuesta con la que examinar la configuración táctica del equipo ha consistido en la observación de las interacciones, combinando el análisis del comportamiento con elementos propios de la sociometría tradicional. Aquí es donde entra el tema principal de este trabajo **el análisis de redes sociales**.

Las redes sociales surgen en el estudio de una variedad de problemas, que van desde

las cuestiones tecnológicas y de transporte a los fenómenos sociales y problemas biológicos, ver Lewis.T.G (2009). Su prevalencia es tal que una rica teoría matemática se ha desarrollado alrededor de ellos.

Mi objetivo en este trabajo es mostrar cómo la teoría matemática de las redes se puede utilizar para analizar la información estadística de los deportes de equipo y medir el rendimiento de un equipo y sus jugadores.

Para ello en primer lugar hablaré de las teorías del análisis de redes y de las diferentes medidas con las que podemos medir la centralidad de los nodos que componen una red social.

A continuación expondré dos ejemplos en los que se han utilizado estas diferentes medidas para medir el rendimiento de un equipo, estos dos ejemplos se harán con los datos de la previa de la final y el partido por el tercer y cuarto puesto del Mundial de Sudáfrica 2010.

Y en el último capítulo del TFG veremos el funcionamiento del análisis de redes en el programa estadístico más utilizado en nuestra carrera, *R*, mediante un ejemplo práctico.

2 ANALISIS DE REDES

“Aquí el más tonto hace relojes de madera y funcionan”

Luis Aragonés (1938-2014)

2.1. Teoría del análisis de redes

En este capítulo se introducirán los conceptos básico de la teoría de redes al igual que las medidas matemáticas obtenidas a partir de estas redes.

En primera instancia se debe tener claro que una **red social** es una estructura social formada por individuos llamados “nodos”, que están ligados por uno o más tipos específicos de interdependencia, como la amistad, el parentesco, el interés común o el intercambio financiero .

El **análisis de redes sociales** considera las relaciones sociales en términos de teoría de redes que consisten en nodos y vínculos (enlaces). Los **nodos** son los actores individuales dentro de las redes y los **vínculos** son las relaciones entre los actores. Las estructuras gráficas resultantes son a menudo muy complejas. Puede haber muchos tipos de vínculos entre los nodos.

Simplificando, una red social es un mapa de lazos especificados (como trayecto del metro, la pertenencia a un grupo social ...) entre los nodos que se están estudiando. Los nodos a los que un individuo está conectado son los **contactos sociales** de ese individuo. Además la red sirve para medir la importancia de un nodo en dicha red social. Estos conceptos se muestran en un diagrama donde los nodos son los puntos y los lazos las líneas.

El análisis de redes ha pasado de ser una metáfora sugerente a un enfoque analítico de una paradigma, con sus enunciados teóricos, métodos, software de análisis de redes sociales e investigadores.

El primer matemático en profundizar en la teoría de redes a partir del problema de los siete puentes de Königsberg fue Euler (1736). La pregunta que se hacían en la época era si era posible cruzar los siete puentes sin tener que pasar por el mismo más de una vez. Si observamos el grafo elaborado por Euler, vemos que esto no es posible.

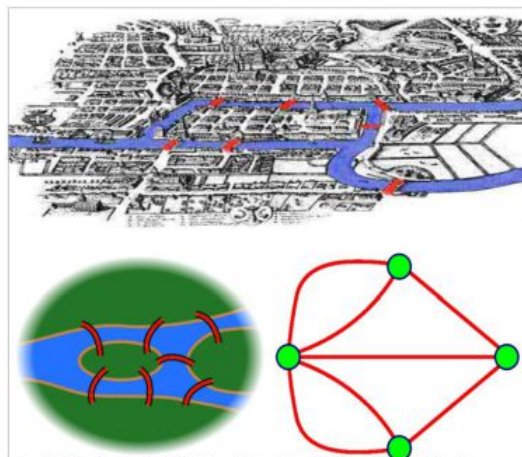


Figura 2.1: Problema de los siete puentes de Königsberg

En 1932, el psiquiatra Jacob Moreno hizo uso de esta misma técnica en una situación bien distinta. En este caso se trataba de una serie de escapadas que se detectaron en el colegio Hudson School for Girls en Nueva York. En menos de dos semanas hasta 14 niñas se escaparon del colegio. Haciendo uso del análisis de redes sociales, Moreno descubrió que las razones que llevaban a las niñas a escapar tenían más que ver con las influencias de su entorno que con la personalidad o la situación personal de las propias niñas, ver Borgatti.P, Mehra.A, Daniel.J, y Labianca.G (2009).

No fue hasta la década de los ochenta cuando el análisis de redes sociales se convirtió en un campo establecido dentro de las ciencias sociales, con una organización profesional, **INSNA**, una conferencia anual, **SUNBELT**, software especializados, **UCINET** e incluso revistas como **Social Networks**.

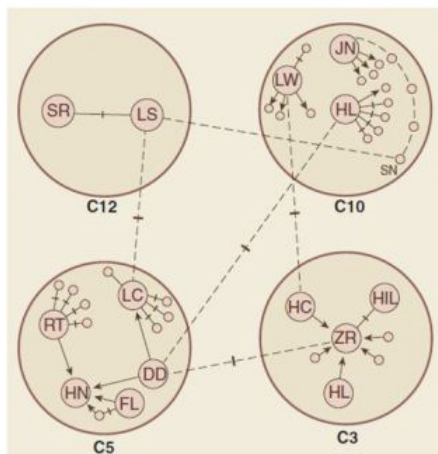


Figura 2.2: Red de Moreno de las relaciones entre las niñas que se escaparon del colegio

Empíricamente el análisis de redes se fundamenta en la creación y desarrollo de la matriz de relaciones y en la construcción del grafo. Cuando va a desarrollarse un análisis relacional, el material básico para el análisis es la construcción de la matriz que liga a los actores entre sí. Frente a la forma tradicional de las variables, los atributos, para llevar a cabo análisis de redes hay que transformar los datos disponibles a una forma relacional, que tiene normalmente la forma de matriz, ver Sanz.L (2003). Si los actores se describen como nodos y sus relaciones como líneas entre pares de nodos, el concepto de red social pasa de ser una metáfora a una herramienta operativa analítica que utiliza el lenguaje matemático de la teoría de grafos, de las matrices y del álgebra relacional.

En el análisis de redes se han desarrollado un gran número de medidas para caracterizar y comparar las estructuras de las redes y las posiciones dentro de ellas. En este

artículo nos centraremos en las medidas de la centralidad de un nodo dentro de un gráfico que determinan la importancia del nodo en una red social (la importancia que tiene una persona dentro de una red social, o la utilidad de un camino dentro de una red urbana).

Hay cinco medidas de centralidad utilizadas ampliamente en el análisis de redes que pasaremos a descubrir a continuación.

2.1.1. Centralidad del grado

La primera y más sencilla es la **centralidad del grado**. Se define como el número de enlaces que inciden sobre un nodo (el número de enlaces que tiene un nodo). Si la red está dirigida, lo que significa que los lazos tienen dirección, entonces usualmente definimos dos medidas separadas de la centralidad del grado: **indegree** y **outdegree**. Indegree es un recuento del número de vínculos dirigidos al nodo y outdegree es el número de vínculos que el nodo dirige a otros.

Esta centralidad de grado se utiliza para calcular **las medidas de Centralización**, que indican la variabilidad de los valores de centralidad entre los nodos del conjunto. Miden el grado en que un nodo tiene una centralidad alta mientras que el resto la tiene baja.

Para un gráfico $G := (V, E)$ con n nodos, V son los vértices o nodos y E los enlaces, el grado de centralidad $C_D(v)$ para el nodo v es:

$$C_D(v) = \frac{\deg(v)}{n - 1} \quad (2.1.1)$$

Siendo $\deg(v)$ el número de enlaces que inciden en v .

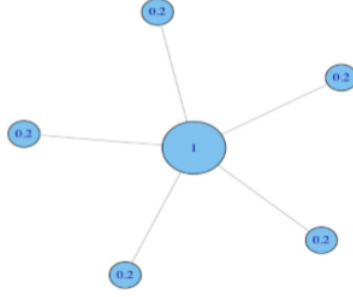


Figura 2.3: Gráfico con valores de centralidad de grado

El cálculo de la centralidad del grado para todos los nodos V de un gráfico $\Theta(E)$ toma una matriz de densidad adyacente que representa al gráfico.

La definición de centralidad puede extenderse a los gráficos, o sea, se puede obtener una medida de centralidad de grado para el gráfico al completo. Sea v^* el nodo con mayor grado de centralidad en el gráfico G . Sea $X := (Y, Z)$, siendo Y el conjunto de nodos y Z las conexiones entre los nodos, el n nodo conectado al gráfico que maximiza la siguiente cantidad (siendo y^* el nodo con mayor grado de centralidad en X):

$$H^* = \max \sum_{j=1}^{|Y|} C_D(y^*) - C_D(y_j) \quad (2.1.2)$$

Entonces el **grado de centralidad** se define:

$$H = \frac{\sum_{i=1}^{|V|} C_D(v^*) - C_D(v_i)}{H^*} \quad (2.1.3)$$

Donde $C_D(v^*)$ es el valor máximo de centralidad en la red y $C_D(v_i)$ es el valor de centralidad para el nodo i . El índice H se define en $[0,1]$. Toma valor 0 cuando todos los nodos tienen exactamente el mismo valor de Centralidad $C_D(v^*)$ y 1 cuando un actor

tiene una importancia predominante sobre el resto.

H se maximiza cuando el gráfico X contiene un nodo que está conectado a todos los nodos y todos los demás están conectados sólo a este nodo central(un gráfico de estrella). En este caso:

$$H = (n - 1)\left(1 - \frac{1}{n - 1}\right) = n - 2 \quad (2.1.4)$$

por lo que el grado de centralidad de G se reduce a:

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v^*) - C_D(v_i)]}{n - 2} \quad (2.1.5)$$

En definitiva la centralidad de grado es una medida eficaz para evaluar la centralidad de un actor en una red social pero desde una perspectiva muy local, ya que mide la importancia y la influencia de dicho nodo con respecto a sus vecinos más cercanos. Por lo tanto para redes sociales con muchos nodos no es una medida óptima para saber la importancia de un nodo en esa red social.

2.1.2. Intermediación

La **intermediación** es una medida de un nodo dentro de un gráfico. Los nodos que se producen en muchos caminos mas cortos entre otros nodos tienen mayor intermediación que aquellos que no.

Para un gráfico $G := (V, E)$ con n nodos, la intermediación se calcula de la siguiente manera :

1. Para cada par de nodos (s, t) calcular todos los caminos mas cortos entre ellos.

2. Para cada par de nodos (s, t) , determine la fracción de caminos más cortos que pasan por el nodo en cuestión.
3. Sumar esta fracción sobre todos los pares (s, t) .

En definitiva según Shivaram.N (2005) :

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (2.1.6)$$

Donde σ_{st} es el número de caminos de s a t y $\sigma_{st}(v)$ es el número de caminos de s a t que pasan por el nodo v . Esto puede normalizarse dividiendo por el número de pares de nodos que no incluyen en v , que es $(n-1)(n-2)$ para grafos dirigidos y $(n-1)(n-2)/2$ para grafos no dirigidos.

Por ejemplo, en un gráfico de estrella unidireccional donde el nodo central, que está contenido en cada trayecto más corto posible, tendrá una intermediación de 1 si se normaliza, mientras que las hojas de las estrellas, que no están en el camino mas corto tendrían una intermediación de 0.

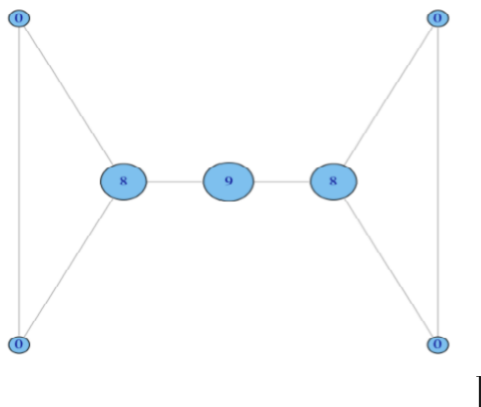


Figura 2.4: Gráfico con valores de intermediación

El cálculo de la intermediación de todos los nodo del gráfico implica calcular los trayectos más cortos entre todos los pares de nodos de un gráfico.

Esto lleva $\Theta(V^3)$, siendo V el conjunto de los nodos, tiempo con el algoritmo de Floyd-Warshall, el cual no encuentra uno sino todos los caminos más cortos entre los dos nodos. En un gráfico escaso de nodos, el algoritmo de Johnson puede ser mas eficiente aunque tarde $\Theta(V^2 \log V + VE)$, donde V sigue siendo en conjuntón de nodos y E el conjunto de lazos, mas tiempo. En las gráficas no ponderadas utilizaríamos el algoritmo de Brandes.U (2000) que tardaría $\Theta(VE)$ para calcular la intermediación.

2.1.3. Cercanía

En topología y áreas relacionadas dentro de las matemáticas, la **cercanía** es uno de los conceptos básicos en un espacio topológico.

El concepto puede definirse naturalmente en un espacio métrico donde se define una noción de distancia entre elementos del espacio, pero también puede ser en espacios topológicos generalizados donde no tenemos una forma concreta de medir distancias.

En la teoría de grafos la centralidad es una medida del nodo dentro de un gráfico. En el análisis de redes la cercanía indica la longitud de la trayectoria mas corta, pues da valores altos a los nodos mas centrados y por tanto se asocia positivamente con otras medidas como la centralidad de grado.

En la teoría del análisis de redes la cercanía es una medida sofisticada de la centralidad. Se define como la distancia geodésica media , la trayectoria más corta entre un



Figura 2.5: Gráfico con valores de cercanía

nodo y todos los otros nodos alcanzables desde él:

$$\frac{\sum_{i \in V} d_G(v, t)}{n - 1} \quad (2.1.7)$$

donde n es el tamaño del componente de conectividad, V todos los enlaces del gráfico, y $d_G(v, t)$ es la distancia mas corta entre el nodo v y el nodo t . La cercanía puede tomarse como una medida de cuanto tiempo tomará la información para extenderse desde un nodo dado a otros nodos alcanzables de la red Newman.M.E.J (2003).

La cercanía $C_C(v)$ de un nodo v es la inversa de la suma de distancias geodésicas a todos los otros nodos de V :

$$C_C(v) = \frac{1}{\sum_{t \in V} d_G(v, t)} \quad (2.1.8)$$

Diferentes métodos y algoritmos pueden ser introducidas en la medida de cercanía, como la centralidad random-walk" (andar aleatorio) introducida por Noh.J.D y Rieger.H (2004) que es una medida de velocidad con la que los mensajes alcanzan al azar un nodo de

otra parte de la red.

Dangalchev.Ch (2006), con el fin de medir la vulnerabilidad de la red, modifica la definición de cercanía para que pueda usarse en gráficos no dirigidos y así la cercanía total sea más fácil de calcular.

$$C_G(v) = \sum_{t \in V} 2^{-d_G(v,t)} \quad (2.1.9)$$

2.1.4. Pagerank

Pagerank centrality, o centralidad de Pagerank, introducida en Brin and L. Page (1998), es una noción de popularidad o importancia la cual sostiene que un nodo es popular si muchos enlaces de otros nodos se dirigen a él. Hay que destacar que esta herramienta fue desarrollada por Google para medir la "popularidad" de una página web, basándose en la cantidad y "calidad" de los enlaces que apuntan a dicha página.

Matemáticamente, la centralidad de Pagerank se define:

$$x_i = p \sum_{j \neq i} \frac{A_{ij}}{L_j^{out}} + (1 - p) \quad (2.1.10)$$

donde L_j^{out} es el número total de enlaces que salen del nodo j al resto del grafo, A_{ji} es el número total de direcciones que salen del nodo j al nodo i , p es un parámetro heurístico que representa la probabilidad de que la "información" de un nodo salga a otros nodos o no y $(1 - p)$ es un parámetro que otorga una popularidad libre a cada nodo.

Hay que tener en cuenta que la puntuación de pagerank de un nodo depende de las

puntuaciones de todos los nodo de de la red. Como consecuencia todas las puntuaciones de pagerank se deben de calcular al mismo tiempo para todos los nodos de la red.

En definitiva, la centralidad de Pagerank asigna, aproximadamente, a cada nodo la probabilidad de que la información llegué a él a través de un razonado intercambio de información en la red mediante los nodos. Si se requiere una precisión adicional para calcular esta probabilidad, la probabilidad p se puede sustituir por las probabilidades dependientes p_i .

2.1.5. Clustering y comunalidades

Un aspecto interesante en el análisis de redes es la forma en la que los nodos intercalan entre ellos. La noción de clustering, o agrupación, nos dice que es una medida del grado en que los nodos tienden a agruparse.

No es una medida de centralidad como tal, pero es una medida cuanto menos interesante que sirve para observar el comportamiento de una red social.

Watts.D y Strogatz.S (1998) propusieron una medida para calcular el clustering local. El clustering local mide el grado en el que los vecinos de un nodo están conectados.

Concretando, el coeficiente de clustering medio es una medida global que indica la probabilidad de que dos vecinos de un nodo de la red escogidos aleatoriamente estén conectados entre sí. Por lo que para un nodo de grado k_i el clustering local se expresa como:

$$c_i^w = \frac{2L_i}{k_i(k_i - 1)} \quad (2.1.11)$$

donde L_i representa en número de lazos entre los vecinos de i . $c_i^w = 0$ implica que ninguno de los vecinos de i tienen conexiones unos con otros, mientras que $c_i^w = 1$

implica que todos los vecinos de i forman un gráfico completo donde todos vinculan entre sí.

Por lo tanto el coeficiente de clustering medio se puede calcular como la media de c_i^w con todos los nodos:

$$c^w = \frac{1}{N} \sum_{i=1}^N c_i^w \quad (2.1.12)$$

El valor c_i^w es la probabilidad con la que dos vecinos de un nodo aleatoriamente seleccionado estén unidos entre ellos.

3 APLICACION DE REDES EN FUTBOL

*“Sin la ayuda de mis compañeros no sería nada de nada.
No ganaría títulos, ni premios ni nada”*

Lionel Messi (1987-)

3.1. Aplicaciones del análisis de redes en el fútbol

El análisis de redes sociales ha surgido como una técnica clave en la sociología moderna. También ha ganado un significativo seguimiento en antropología , biología, economía, geografía, psicología social., y más recientemente en el deporte, específicamente en el fútbol 11.

En este campo la red social de un equipo pasará a ser una red de pase, la cual contiene a los jugadores de un equipo como nodos y los arcos de conexión entre dos jugadores estarán ponderados por el número exitoso de pases completados entre ellos y esto proporciona una visión inmediata de la táctica que un equipo utiliza , si un equipo tiene a abusar de pases largos o cortos o si un jugador no interviene demasiado en el juego del equipo...

Cierto es que la red de pase es estática y los jugadores no se mantienen los 90 minutos sin moverse de esa posición, pero esta manera de representación nos permite examinar el papel de los jugadores individualmente así como al equipo.

En este apartado se verá la utilización de las técnicas métricas en ejemplos concretos.

3.1.1. Caso I: Previa a la final de Sudáfrica. España vs Holanda

En este capítulo se presentan los resultados que se dieron al computar diferentes medidas de centralidad, las mencionadas en el capítulo anterior, en los finalistas de final del mundial de la FIFA 2010, España y Holanda.

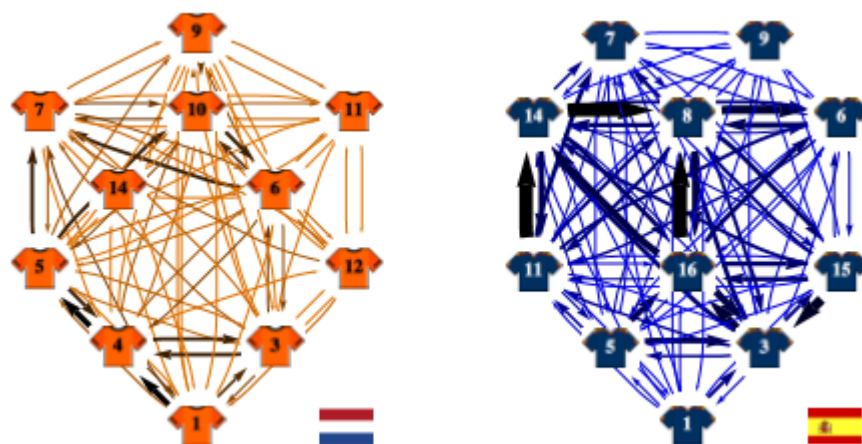


Figura 3.1: Red de pases de las selecciones holandesas y españolas

Los datos de pases del partido de la final fueron descargados de la página web de la FIFA utilizando un script Python personalizado por los redactores del artículo "A network theory analysis of football strategies", Saramaki.J, Kivel.M, Onnela.J.P, Kaski.K, y Kertesz.J (2007), de donde se han extraído los datos.

Las tablas 1 y 2 muestran las puntuaciones de la cercanía, la intermediación, el pagerank y el clustering, respectivamente, de los jugadores que jugaron la final.

La conclusión a la que se puede llegar tras ver los datos es que hay una alta correlación entre las puntuaciones mas altas de cercanía, pagerank y clustering, que confirman la percepción que tuvieron los medios del rendimiento empleado por los jugadores, un

Jugador	$C_C(v)$	$C_B(v)$	x_i	c_i^w
Casillas	16.52	0.00	3.29	20.46
Pique	17.32	3.92	11.46	30.70
Puyol	16.32	2.86	7.92	27.12
Iniesta	14.6	0.50	8.54	31.03
Villa	8.68	0.50	5.89	23.96
Xavi	18.28	1.19	14.66	46.47
Capdevila	16.54	6.12	10.56	29.1
Alonso	17.11	1.19	12.31	41.69
Ramos	16.45	2.41	9.02	27.05
Busquets	18.55	2.41	12.99	35.32
Pedro	3.42	0.00	3.55	16.75

Tabla 3.1: Puntuaciones de los jugadores de la selección de España, Mundial 2010.

Jugador	$C_C(v)$	$C_B(v)$	x_i	c_i^w
Stekelenburg	16.34	0.32	7.63	28.35
Van Der Wiel	14.43	2.97	9.79	31.39
Heitinga	16.23	2.67	11.06	31.34
Mathijsen	17.30	1.30	10.84	33.22
V.Bronckhorst	15.74	1.12	10.07	37.00
Van Bomme	12.46	3.08	11.19	32.36
Kuyt	7.97	1.67	9.02	27.06
De Jong	10.95	2.73	9.28	28.36
Van Persie	6.89	2.92	5.88	20.13
Sneijder	10.91	2.17	10.32	33.77
Robben	5.91	0.16	4.91	23.91

Tabla 3.2: Puntuaciones de los jugadores de la selección de Holanda, Mundial 2010.

ejemplo es la alta puntuación en estas dos medidas de Xavi, posiblemente el jugador mas importante de la Selección Española.

En el lado español, cabe destacar que las puntuaciones de intermediación son bajas y uniformemente distribuidas, la defensa y el centrocampista defensivo y en particular Pique y Capdevila, son los que más alto tienen el valor de intermediación debido a que eran los principales iniciadores del juego español, esto es signo de que España posee una estrategia de paso bien balanceada y las puntuaciones altas de clustering demuestran que la selección española es un equipo extremadamente bien conectado en el que casi todos los jugadores se ayudan ofreciéndose como opción de pase. La única excepción es Pedro, cuyas puntuaciones bajas se pueden dar a la posición que tenía en el campo, jugó de delantero, y además no jugaba los 90 minutos. Cabe destacar que los delanteros pueden ser identificados como aquellos jugadores que tienen la menor cercanía, intermediación y pagerank(vease David Villa), ya que son jugadores que están aislados de la construcción de juego y están esperando a recibir el pase final, lo mismo pasa con los jugadores que son reemplazados mas a menudo.

Las puntuaciones holandesas están en general cercanas a las de las españolas, lo que indica el buen juego realizado por los holandeses en su camino a la final, pero hay algunas diferencias notables. En primer lugar hay una clara diferencia en la densidad de pases que se observan en la figura 1.

En segundo lugar, se observa por la cercanía $C_C(v)$ los jugadores holandeses no están tan cerca(conectados por pases cortos) los unos de los otros y tienen el pagerank distribuido mas uniformemente entre sus jugadores. No tienen un jugador que predomine tanto como Xavi en España.

Cabe destacar que en los dos equipos los que mayor valor de clustering tiene son los centrocampistas.

3.1.2. Caso II: Previa al 3° puesto de Sudáfrica. Alemania vs Uruguay

Las tablas 3 y 4 muestran la misma información que las dos anteriores tablas pero esta vez para las selecciones de Alemania y Uruguay respectivamente. Las alineaciones puestas en las tablas son las que usaron los dos equipos en las semifinales.

Jugador	$C_C(v)$	$C_B(v)$	x_i	c_i^w
Neuer	7.58	0.37	4.74	21.54
Friedrich	9.29	3.55	10.08	24.99
Khedira	8.70	10.58	11.38	26.31
Schweinsteiger	10.28	13.17	17.32	27.35
Ozil	7.54	4.34	10.05	22.62
Podolski	4.91	0.22	6.66	30.21
Klose	0.92	0.00	2.48	14.34
Trochowski	3.00	0.00	2.85	33.02
Lahm	10.60	11.83	14.65	24.56
Mertesacker	10.81	3.42	13.27	26.71
Boateng	6.85	3.63	6.25	19.85

Tabla 3.3: Puntuaciones de los jugadores de Alemania, Mundial 2010.

Los resultados de estos dos equipos apuntan a una mayor diferencia en sus conexiones que las mostradas por los finalistas. Alemania con sus puntuaciones de cercanía y clustering mas grandes que las de Uruguay nos muestran que Alemania está en general mejor conectada que su rival. Sin embargo, a diferencia de lo que ocurría en las selecciones finalistas, donde las puntuaciones estaban distribuidas uniformemente, estas dos

Jugador	$C_C(v)$	$C_B(v)$	x_i	c_i^w
Muslera	0.88	1.98	4.62	9.96
Godin	1.80	4.20	8.37	11.99
Gargano	0.76	0.37	2.98	7.33
Victorino	1.75	0.88	5.59	14.28
Cavani	1.61	10.22	10.22	13.68
Forlán	2.08	10.29	13.12	15.02
A.Pereira	1.90	3.75	9.01	16.45
Perez	2.36	10.63	15.25	19.12
M.Pereira	2.28	1.51	12.21	20.07
Arevalo	2.45	5.85	13.12	19.83
Caceres	1.34	3.65	5.52	9.51

Tabla 3.4: Puntuaciones de los jugadores de Uruguay, Mundial 2010.

selecciones dependen mas del esfuerzo de algunos de sus jugadores para hacer circular el balón, en Alemania por ejemplo Lahm y Schweinsteiger desempeñan un papel central en su equipo, ya que son los que mas pases reciben(Page rank más alto) y los jugadores que más conectan con pases cortos con el resto de compañeros(intermediación).

En el caso de Uruguay se ve que dos de los jugadores con la intermediación más alto lo tienen los delanteros, Forlán y Cavani, esto, al contrario que pasaba con la Selección Española, se debe a que en vez de iniciar el juego desde abajo con pases cortos el juego de Uruguay era mas directo hacia los delanteros, esto también se deduce viendo que Forlán es el segundo jugador de Uruguay mas “popular”(su Page Rank es el 2º mas alto) del equipo .

Otra medida la cual induce a pensar que Uruguay práctica un juego directo es el bajo nivel de cercanía de sus jugadores. Ciertamente es que a pesar de su juego directo los centrocampistas , Perez y Arevalo , tienen un gran peso en la red de pase dado que al ser centrocampistas de corte defensivo eran los encargados de recuperar el balón y una vez recuperado enlazar con los delanteros uruguayos.

4 APLICACION A LA FINAL FIFA

*“El balón está hecho de cuero, el cuero viene de la vaca,
la vaca come pasto, así que hay que echar el balón al pasto”*

Di Stéfano (1926-2014)

4.1. Análisis de redes en RStudio

Si bien la teoría de las redes sociales puede aplicarse fácilmente en la investigación teoría y lo estudios empíricos cualitativos, se hace hincapié en el uso de software para analizar y visualizar los datos de la red una vez que se han recogido. Existen varios paquetes de software disponibles, pero dos paquetes de R (Network e iGraph) se han convertido quizá en las herramientas más flexibles y potentes para realizar análisis de red. Este apartado está dedicado a ver el funcionamiento del paquete iGraph mediante un ejemplo, el cual se trata del partido jugado por las selecciones Españolas y Holandesas en el Mundial de 2014 para así poder comparar los resultados dado por los datos del capítulo 2:

4.1.1. Paquete iGraph

El principal paquete que utilizaré será **iGraph**. iGraph es un paquete de software para gráficos sencillos y análisis de redes. Puede manejar grandes gráficos con facilidad y proporciona funciones para generar gráficos aleatorios y regulares, visualización de gráficos, métodos de centralidad. . . De todas las posibilidades que ofrece el paquete iGraph se tomara en cuenta en este TFG la creación de redes, generar y visualizar gráficos y hallar las medidas de los métodos de centralidad estudiados en este trabajo. Para ello se

utilizará principalmente el paquete iGraph en detrimento del paquete network ya que se trata de un paquete más útil a la hora de medir las distintas medidas de centralidad mencionadas a lo largo del artículo.

4.1.2. Resultados

En este apartado se pretende realizar un análisis comparativo entre los datos de centralidad del capítulo 2 de la previa de la final del Mundial del 2010 y los obtenidos del ejemplo realizado con la herramienta estadística R del partido de Holanda vs España del Mundial 2014. Antes de comparar los datos primero vamos a realizar el análisis de los datos del partido del 2014 entre las dos selecciones. Las siguientes tablas muestran las puntuaciones de la centralidad del grado(indegree), la intermediación, el pagerank y la cercanía, respectivamente, de los jugadores que compusieron los 11 iniciales.

Jugador	$C_D(v)$	$C_B(v)$	x_i	$C_C(v)$
Casillas	15	0.00	3.397 %	0.625
Alba	42	0.0629	8.414 %	0.769
Pique	51	4.952 %	9.599	0.909
Ramos	48	2.206	8.846 %	0.909
Azpilicueta	35	1.646	7.071 %	0.909
Busquets	60	4.106	10.875 %	1
Alonso	62	1.418	11.76 %	0.909
Xavi	60	0.797	11.47 %	0.909
Iniesta	74	2.242	14.268 %	0.909
Silva	47	0.566	9.672 %	0.909
Costa	19	0.00	4.617 %	0.833

Tabla 4.1: Puntuaciones de los jugadores de la selección de España, Mundial 2014.

Antes de realizar el análisis de los datos hay que hacer unas aclaraciones: los resultados del page rank vienen dado en porcentajes ya que lo que nos indica es el porcentaje sobre el total de pases (o iteraciones) que recibió cada jugador a lo largo del partido y

Jugador	$C_D(v)$	$C_B(v)$	x_i	$C_C(v)$
Cillessen	17	0.874	6.312 %	0.769
Vlaar	23	1.420	8.011 %	0.769
De Vrij	22	1.621	7.894 %	0.909
Martins	29	2.212	9.966 %	0.833
Blind	39	3.452	13.371 %	0.909
De Jong	27	4.046	10.745 %	0.909
Janmaat	17	2.354	7.791 %	0.833
De Juzman	20	1.225	7.015 %	0.769
Van Persie	29	0.803	9.405 %	0.909
Sneijder	28	1.223	10.151 %	0.833
Robben	27	0.765	9.334 %	0.909

Tabla 4.2: Puntuaciones de los jugadores de la selección de Holanda, Mundial 2014, Mundial 2014.

en cuanto a los datos de cercanía indicar que los resultados oscilan entre 0 y 1, donde un jugador con una puntuación de 0 indica que no ha intervenido en el partido y 1 un jugador que ha estado en total contacto con el resto de jugadores de una manera influyente.

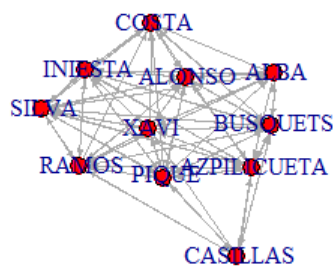


Figura 4.1: Red de pase de la selección española

Como se puede observar en la columna de centralidad de grado vemos que la Selección

Española tuvo mayor número de pases entre sus jugadores que la selección holandesa por lo que tuvo mas posesión de balón, pero esto no quiere decir que llevará más peligro hacia la portería de su rival ya que Van Persie, el delantero centro holandés, fue el segundo receptor de pases de su selección(29) en contra del punta español, Diego Costa(19), que fue el que menos pases recibió después del portero.

En cuanto a la intermediación se puede ver como a diferencia de la Selección Española del mundial del 2010 donde Xavi, a pesar de no ser la puntuación más alta, tenía cierta importancia en la conexión de la jugada vemos que en este partido su puntuación es la mitad de la de 4 años atrás por lo que el cerebro de la selección española no tuvo la misma importancia en este partido. También cabe destacar las puntuaciones de intermediación de los laterales, los cuales en el año 2010 tuvieron una alta puntuación, Capdevila tuvo la máxima(6,12), y sin embargo en este partido los laterales tienen una baja puntuación, sobre todo Jordi Alba(0,0629), lo que indica que la salida por bandas de la Selección Española no fue buena. En cambio los datos de intermediación de la Selección Holandesa a pesar de no ser mucho mas superiores a la de la Selección Española, si que están mucho mejor balanceados entre todos los jugadores no hay ningún jugador con 0 absoluto como ocurre con Casillas y Costa en la selección española. Lo que indica que la Selección Holandesa tuvo una mejor conexión de pase entre todos sus jugadores que la selección Española.

Se puede apreciar que los valores más altos de PageRank de la Selección Española recaen en los centrocampistas, entre Alonso, Xavi, Busquets e Iniesta suman mas de un



Figura 4.2: Red de pase de la selección holandesa

48 % de los pases recibidos, osea que en el juego del combinado español los centrocampistas son los que mas importancia tienen en el juego y por ello son los jugadores que más balones reciben. En cambio en Holanda los valores están más equilibrados, excepto por el Lateral izquierdo Blind, el cual recibe más del 13 % de los balones, además su Delantero centro ,Van Persie, tiene un 9,4 de PageRank mientras el delantero centro español Diego Costa ni siquiera llega al 5 por lo que la delantera holandesa tuvo más peso en el juego que la española.

En definitiva podemos decir que a pesar de que la Selección Española controló durante más tiempo el balón sus transiciones hacia el ataque fueron ineficientes y las combinaciones se realizaban entre centrocampistas y defensas centrales sin apenas tener salida por banda ni dar balones a la delantera sin embargo la Selección Holandesa a pesar de dar menos pases entre sus jugadores sus jugadas fueron más verticales y mejor trenzadas con la participación de todo el equipo sabiendo hacer llegar el balón a su delantera. Por

lo tanto la importancia de los pases en un equipo de fútbol no reside en el número total de pases que den sino en la buena conexión que halla entre sus jugadores como bien refleja el marcador de este partido donde la Selección Holandesa avasalló a la Española y le endosó un 5 a 1.


```
#####
# {Cargar librerías:}
# {library(readr)}
#####
library(readr)
library(igraph)
library(network)

matrizespaña<- read_delim("~/Tfg r/matrizconpesosespaña.csv",
";", escape_double = FALSE, col_names = FALSE,
na = "empty")

colnames(matrizespaña)<- c("CASILLAS","ALBA","PIQUE","RAMOS",
"AZPILICUETA","BUSQUETS","ALONSO","XAVI",
"INIESTA","SILVA","COSTA")matrizespaña<- read_delim("~/Tfg r/matrizconpesosespaña.csv",
";", escape_double = FALSE, col_names = FALSE,
na = "empty")

colnames(matrizespaña)<- c("CASILLAS","ALBA","PIQUE","RAMOS",
"AZPILICUETA","BUSQUETS","ALONSO","XAVI",
"INIESTA","SILVA","COSTA")

linksespaña= (matrizespaña)linksespaña= (matrizespaña)

nodosespaña <- read_delim("~/Tfg r/nodosespaña.csv",
";", escape_double = FALSE, trim_ws = TRUE)
linksespaña =as.matrix(linksespaña,dimnames=list(C))

red1= graph.adjacency(linksespaña,mode= "directed")

red1b <- simplify(red1,remove.multiple = T, remove.loops = F,
edge.attr.comb=c(weight="sum", type="ignore"))
plot(red1b, vertex.label.dist=.5,edge.arrow.size= 0.3,vertex.color= "red")

matrizholanda<- read_delim("~/Tfg r/matrizconpesosholanda.csv",
";", escape_double = FALSE, col_names = FALSE,
na = "empty")
colnames(matrizholanda)<- c("CILLESEN","VLAAR","DE VRIJ","MARTINS",
"BLIND","DE JONG","JANMAAT","DE GUZMAN",
"VAN PERSIE","SNEIJDER","ROBBEN")

linksholanda= (matrizholanda)

nodosholanda <- read_delim("~/Tfg r/nodosholanda.csv",
";", escape_double = FALSE, trim_ws = TRUE)
linksholanda =as.matrix(linksholanda,dimnames=list(C))

red2= graph.adjacency(linksholanda,mode= "directed")

red2b <- simplify(red2,remove.multiple = T, remove.loops = F,
edge.attr.comb=c(weight="sum", type="ignore"))
plot(red2b, vertex.label.dist=.5,edge.arrow.size= 0.3,vertex.color= "orange")
```

```
centr_degree(red1,mode= "out",normalized=T)
centr_degree(red1,mode= "in",normalized=T)
centr_betw(red1,directed= TRUE,normalized=T)
page_rank(red1,algo =c("prpack"),directed = TRUE)
resultados= (page_rank(red1,algo =c("prpack"),directed = TRUE)$vector)*100
centr_clo(red1,mode="in",normalized =T)

centr_degree(red2,mode= "out",normalized=T)
centr_degree(red2,mode= "in",normalized=T)
centr_betw(red2,directed= TRUE,normalized=T)
page_rank(red2,algo =c("prpack"),directed = TRUE)
resultados2= (page_rank(red2,algo =c("prpack"),directed = TRUE)$vector)*100
centr_clo(red2,"in",normalized =T)
```

- Borgatti.P, Mehra.A, Daniel.J, y Labianca.G. (2009). Network analysis in the social sciences. , 2–3.
- Brandes.U. (2000). A faster aloritm for betweenness centrality.
- Brinandl.S, y Page.L. (1998). The anatomy of a large-scale hypertextual web search engine. , 107–117.
- Castellano, y Blanco, H. (2002). *Analisis diacronico de la acci juego en futbol*.
- Chicote.J, Morante.J.C, y Vaquera.A. (2009). Propuesta metodola para el ansis tico de las acciones ofensivas en equipos profesionales de baloncesto.
- Dangalchev.Ch. (2006). Residual closeness in networks, phisica. , 365–556.
- Euler, L. (1736). *Solutio problematis ad geometriam situs pertinentis*.
- Lago.C, y Anguera.M.T. (2003). Utilizacil ansis secuencial en el estudio de las interacciones entre jugadores en el ftbol de rendimiento.
- Lewis.T.G. (2009). Network science: Theory and practice.
- Newman.M.E.J. (2003). Disordered systems and neural networks.
- Noh.J.D, y Rieger.H. (2004). Phys. rev. lett. 92.
- Sanz.L. (2003). Ansis de redes sociales: o cepresentar las estructuras sociales subyacentes.
- Saramaki.J, Kivel.M, Onnela.J.P, Kaski.K, y Kertesz.J. (2007). Generalizations of the clustering coefficient to weighted comple netwoks.
- Shivaram.N. (2005). The betweenness centrality of biological networks.
- Watts.D, y Strogatz.S. (1998). Collective dynamics of 'small-world' networks.